Almost Independent and Weakly Biased Arrays: Efficient Constructions and Cryptologic Applications

Jürgen Bierbrauer[1] and Holger Schellwat[2]

[1] Department of Mathematical Sciences, Michigan Technological University, Houghton, Michigan 49931, USA
jbierbra@mtu.edu
[2] Department of Natural Sciences, Örebro University, SE-70182 Örebro, Sweden
holger.schellwat@nat.oru.se

Abstract. The best known constructions for arrays with low bias are those from [1] and the exponential sum method based on the Weil-Carlitz-Uchiyama bound. They all yield essentially the same parameters. We present new efficient coding-theoretic constructions, which allow far-reaching generalizations and improvements. The classical constructions can be described as making use of Reed-Solomon codes. Our recursive construction yields greatly improved parameters even when applied to Reed-Solomon codes. Use of algebraic-geometric codes leads to even better results, which are optimal in an asymptotic sense. The applications comprise universal hashing, authentication, resilient functions and pseudorandomness.

Key Words

Low bias, almost independent arrays, Reed-Solomon codes, Hermitian codes, Suzuki codes, Fourier transform, Weil-Carlitz-Uchiyama bound, exponential sum method, Zyablov bound, hashing, authentication, resiliency.

1 Introduction

The concepts of limited dependence and low bias have manifold applications in cryptography and complexity theory. We mention universal hashing, authentication, resiliency against correlation attacks, pseudorandomness, block ciphers, derandomization, two-point based sampling, zero-knowledge, span programs, testing of combinatorial circuits, intersecting codes, oblivious transfer, interactive proof systems, resiliency (see [19, 16, 18, 17, 1, 11, 25, 10, 6, 9, 7, 13, 16]). A basic notion underlying these concepts are families of $\epsilon-$biased random variables. The Weil-Carlitz-Uchiyama bound and several constructions from the influential papers by Naor and Naor [18] and by Alon, Goldreich, Håstad and Peralta [1] provide families of $\epsilon-$biased random variables. All these classical constructions yield very similar parameters. In this paper we describe methods, which generalize these constructions and yield far-reaching improvements. Essential ingredients are linear codes and the Fourier transform.

2 Bias and Dependency

We use neutral notation which is suited to describe all the applications (hashing, authentication, derandomization, pseudorandomness, ...).

Definition 1. *Let p be a prime. An $(n,k)_p-$array $\mathcal{A}$ is an array with n rows and k columns, where the entries are taken from a set with p elements.*

Definition 2. *Let p be a prime, $v=(v_1,v_2,\ldots,v_n)\in \mathbb{F}_p^n$. For every $i\in\mathbb{F}_p=\mathbb{Z}/p\mathbb{Z}$ let $\nu_i(v)$ be the frequency of i as an entry of v. Let ζ be a primitive complex $p-$th root of unity. The* **bias** *of v is defined as*

$$bias(v)=\frac{1}{n}\left|\sum_{i\in\mathbb{F}_p}\nu_i(v)\zeta^i\right|$$

We have $0\leq bias(v)\leq 1$. As $\sum_{i\in\mathbb{F}_p}\zeta^i=0$ the bias is low if all elements of $\mathbb{F}_p$ occur with approximately the same frequency as entries in v.

Definition 3. *Let $0\leq\epsilon<1$. An $(n,k)_p-$array is $\epsilon-$***biased** *if every nontrivial linear combination of its columns has bias $\leq\epsilon$.*

The bias of an array is a property of the $\mathbb{F}_p-$linear code generated by the columns. The bias of the array is low if and only if every nonzero word of the code has low bias.

While the bias of a vector depends on the choice of the root of unity, the bias of an array is independent of this choice.

Definition 4. *Let $0\leq\epsilon<1$. An $(n,k)_p-$array is* **t-wise $\epsilon-$biased** *if every nontrivial linear combination of at most t of its columns has bias $\leq\epsilon$.*

Definition 5. *Let $0\leq\epsilon<1$. An $(n,k)_p-$array $\mathcal{A}$ is* **t-wise ϵ -dependent** *if for every set U of $s\leq t$ columns and every $a\in\mathbb{F}_p^s$ the frequency $\nu_U(a)$ of rows of $\mathcal{A}$, whose projection onto U equals a satisfies*

$$\left|\frac{\nu_U(a)}{n}-1/p^s\right|\leq\epsilon.$$

The notion of a t-wise ϵ -dependent array generalizes the combinatorial notion of an orthogonal array of strength t (equivalently: $t-$universal family of hash functions in the sense of Carter/Wegman [11]). An array is t-wise independent (=0-dependent) if and only if it is an orthogonal array of strength t.

The most important of these concepts from the point of view of applications is t-wise ϵ -dependency. It captures the familiar theme of representing a family of random variables (the columns of the array) on a small sample space (the rows of the array, with uniform distribution) such that any t of the random variables are almost statistically independent.

We want to point out in the sequel that the construction problem of t-wise ϵ -dependent arrays can be efficiently reduced to the construction of $\epsilon-$biased arrays. This is the basic idea behind [18].

The following construction of t-wise $\epsilon-$biased arrays is essentially from [18].

Theorem 1. *Let the following be given:*

- *An* $(n, k)_p-$*array* B, *which is* $\epsilon-$*biased.*
- *A linear code* $[N, N-k, t+1]_p$.

Then we can construct an $(n, N)_p-$*array, which is* t*-wise* $\epsilon-$*biased.*

Theorem 2. *An array, which is* t*-wise* $\epsilon-$*biased, is also* t*-wise* $\epsilon'-$*dependent for some* $\epsilon' < \epsilon$.

The fundamental Theorem 2 is proved in a nontrivial but standard way by using the Fourier transform, see [5]. The following construction from the journal version of [16] is obvious and useful:

Theorem 3. *If there is an array* $(n, k')_p$, *which is* $t'-$*wise* $\epsilon-$*dependent, and* $t \leq t'/l,\ k \leq k'/l$, *then there is an array* $(n, k)_{p^l}$, *which is* $t-$*wise* $\epsilon-$*dependent.*

We see that indeed the central problem is to efficiently construct $\epsilon-$biased arrays. Linear codes are then used to construct $t'-$wise $\epsilon-$biased arrays via Theorem 1. The standard method is to use BCH codes. The resulting $t'-$wise $\epsilon-$biased arrays are also $t'-$wise $\epsilon-$dependent by Theorem 2. Because of Theorem 3 it is possible to concentrate entirely on binary arrays. We turn to the basic problem of constructing weakly biased arrays.

Definition 6. *Denote by* $f_p(b, e)$ *the minimum* a *such that there is an array* $(p^a, p^b)_p$, *which is* $p^{-e}-$*biased.*

Clearly $f_p(b, e)$ is weakly monotonely increasing in both arguments. The construction from [1] shows the following:

Theorem 4. *There is an efficient construction showing*

$$f_p(b, e) \leq 2(b + e).$$

3 The Weil-Carlitz-Uchiyama Construction

The celebrated Weil-Carlitz-Uchiyama bound [8] may be understood as a limit on the bias of dual BCH-codes. More precisely, let (a_j) be a basis of $\mathbb{F}_{p^f}|\mathbb{F}_p$ and $Tr : \mathbb{F}_{p^f} \longrightarrow \mathbb{F}_p$ the trace. Consider the array $\mathcal{A}$ whose rows are indexed by the elements $\alpha \in \mathbb{F}_{p^f}$ and whose columns are indexed by $a_j X^i$, where $i \leq n$ and i is not a multiple of p. The corresponding entry is $Tr(a_j \alpha^i)$. The WCU bound asserts that this $(p^f, f(n - \lfloor n/p \rfloor))_p-$array has bias $\leq (n-1)p^{-f/2}$.

Comparison reveals that the WCU construction (exponential sum method) yields parameters which are very similar to (a little better than) Theorem 4. All constructions based on one of these classical methods will produce about the same parameters.

4 The Zyablov Construction

As remarked earlier Theorem 3 makes it possible to base the construction on **binary** $\epsilon-$biased arrays. This has the advantage that a direct link to coding theory can be used. An array $(n,k)_2$ is $\epsilon-$biased if and only if the code generated by the columns has dimension k and the relative weights of all nonzero codewords are in the interval of length ϵ centered at $1/2$. This elementary observation yields an immediate reduction of the construction problem of binary weakly biased arrays to the construction of linear codes containing the all-1-vector.

Theorem 5. *Let $0 \leq \epsilon < 1$. The following are equivalent:*

- *An $(n,k)_2$-array, which is $\epsilon-$biased.*
- *A binary linear code of length n and dimension $k+1$, which contains* $\mathbf{1}$ *and whose minimal distance d satisfies*

$$\frac{d}{n} \geq \frac{1-\epsilon}{2}$$

Constructing families of $\epsilon-$biased $(n,k)_2-$arrays which are asymptotically nontrivial (meaning that ϵ is fixed and $k/n \geq R > 0$) is equivalent to constructing asymptotically nontrivial families of binary linear codes containing the all-1-vector.

The question of determining the asymptotics of binary codes is one of the most famous and most well-studied problems in coding theory. The question is how incisive the additional condition is. A famous simple result is the **Gilbert-Varshamov bound:** for every prime-power q and $\delta < (q-1)/q$ the rate $R = 1-H_q(\delta)$ can be asymptotically reached by families of $q-$ary linear codes. It can be managed that the all-1-word is contained in all these codes. Unfortunately this bound is not constructive.

The construction given in [15] does not yield linear codes. The **Justesen-method** [14, 21] is constructive, but the all-1-word is not contained in the resulting codes. The Justesen method when applied to families of algebraic-geometric codes yields precisely the **Zyablov bound.** However, for the same reason as above this does not yield families of binary $\epsilon-$biased arrays.

More interesting for our problem is the original semi-constructive proof of the Zyablov bound [27]. In fact, apply concatenation to a Reed-Solomon code $[q^m, rq^m, (1-r)q^m]_{q^m}$ as outer code and a code $[n,m,d]_q$ as inner code, where it is assumed that the inner code asymptotically meets the Gilbert-Varshamov bound $(d/n = \mu, m/n = 1-H_q(\mu))$. The concatenated code has parameters $[q^m n, rq^m m, (1-r)q^m d]_q$, with relative distance $\delta = (1-r)\mu$ and rate $R = r(1-H_q(\mu))$. This construction shows that for every $\mu < (q-1)/q$ and $\delta < \mu$ we can construct families of $q-$ary linear codes with relative distance δ and rate $R \geq (1-H_q(\mu)(1-\delta/\mu)$. The only drawback is that this is not really constructive. However, for short inner codes this may be feasible. Let us explore the situation in more detail.

We aim at a lower bound for $f_2(b,e)$. Choose $r = 2^{-(e+1)}, \mu = \frac{1}{2} - 2^{-(e+2)}$. As the relative distance of the concatenated code is $(1-r)\mu$ we obtain as bias

$\epsilon = 1-2(1-r)\mu = 1-(1-2^{-(e+1)})^2 = 2^{-e}-2^{-(2e+2)}$. It follows that $\epsilon < 2^{-e}$. We have $b = m+\log(m)-e-1$ and $a = m+\log(n)$. What is the order of magnitude of the rate $S = 1 - H_2(\mu)$ of the inner code guaranteed by Gilbert-Varshamov?

We have $\mu = \frac{1}{2} - 2^{-(e+2)} = (2^{e+1}-1)/2^{e+2}, 1-\mu = (2^{e+1}+1)/2^{e+2}$ and $S = 1-2^{-(e+2)}((2^{e+1}-1)(e+2-\log(2^{e+1}-1))+(2^{e+1}+1)(e+2-\log(2^{e+1}+1))$. Collecting the terms without log yields $S = 2^{-(e+2)}((2^{e+1}-1)\log(2^{e+1}-1) + (2^{e+1}+1)\log(2^{e+1}+1)) - (e+1)$. Divide the arguments of the log $-$terms by 2^{e+1}. The term obtained from compensating for that is $e+1$ and cancels against the last summand. We obtain $S = 2^{-(e+2)}((2^{e+1}-1)\log(1-2^{-(e+1)}) + (2^{e+1}+1)\log(1+2^{-(e+1)}))$. Using the series for $\ln(1 \pm x)$ we obtain

$$S = \frac{2^{-(e+2)}}{\ln(2)}(((2^{e+1}-1)(-2^{-(e+1)}-2^{-2e-3}-\ldots)+(2^{e+1}+1)(2^{-(e+1)}-2^{-2e-3}+)$$

$$= 2^{-(e+2)}\frac{1}{\ln(2)}(-1+2^{-(e+1)}-2^{-e-2}\ldots+1+2^{-(e+1)}-2^{-e-2}\ldots),$$

where terms involving $-2e$ in the exponent and higher have been omitted. This yields $S \sim 2^{-(2e+3)}/\ln(2)$.

Theorem 6. *The Zyablov method needs the construction of binary $[n,m,d]_2$ codes, where $n = \ln(2)2^{2e+3}m$ and $d/n = \frac{1}{2} - 2^{-(e+2)}$. The output is a weakly biased array showing*

$$f_2(m+\log(m)-e-1, e) \leq m+\log(m)+2e+3.$$

Theorem 6 states $f_2(b,e) \leq b+3e+4$. It improves on the bound from Theorem 4 when $b > e$.

5 A Coding-Theoretic Construction of Weakly Biased Arrays

In Section 4 we used an equivalent coding-theoretic interpretation of binary weakly biased arrays to obtain constructions. Observe however that this does not seem to lead to explicit asymptotic constructions. The Zyablov method presupposes exhaustive search for codes of moderate length attaining the Gilbert-Varshamov bound.

When $p > 2$ an equivalent reduction to coding theory is not available. Our next theorem provides a general link, which allows the use of linear codes in the construction of $p-$ary weakly biased arrays. As this leads to efficient constructions, it is interesting even in the binary case.

Theorem 7. *Let $\mathcal{C}$ be a code $[n,k,d]_q$, where $q = p^m$ and $\mathcal{B}$ an $(n_0,m)_p$-array of bias ϵ_0. We can construct an $(nn_0,km)_p-$array with bias $\epsilon = 1-\delta+\delta\epsilon_0 < 1-\delta+\epsilon_0$, where $\delta = d/n$ is the relative distance of code $\mathcal{C}$.*

A proof of Theorem 7 is in [5]. Application of Theorem 7 to Reed-Solomon codes $[p^m, Rp^m, (1-R)p^m]_{p^m}$ and inner unbiased arrays $(p^m, m)_p$ (consisting of all m-tuples) yields the following:

Theorem 8. *For every natural number m and every rational number $0 < \epsilon_0 < 1$ with denominator p^m we can construct an array $(p^{2m}, m\epsilon_0 p^m)_p$ with bias $\leq \epsilon_0$.*

In particular Theorem 8 yields yet another proof for the parameters from Theorem 4 and from the WCU construction.

Our Theorem 7 is much more general. In order to obtain essential improvements on Theorem 4 let us consider a recursive application. Apply Theorem 7 with a Reed-Solomon code $[p^m, Rp^m, (1-R)p^m]_{p^m}$, where $R = \epsilon/2$ and an $\epsilon/2-$biased $(4m^2/\epsilon^2, m)_p-$array. We obtain the following:

Theorem 9. *We can construct arrays $(4m^2p^m/\epsilon^2, m\epsilon p^m/2)_p$, which are $\epsilon-$biased. The choice $m = p^j, \epsilon = p^{-e}$ yields $f_p(p^j + j - e - 1, e) \leq p^j + 2j + 2e + 2$.*

Theorem 9 states in particular $f_p(b, e) \leq b + 3e + j + 3$, where $j \sim \log(b + e)$. In the binary case this is very close to Theorem 6 and it yields an essential improvement over Theorem 4 when $b > e$.

Example 1. Apply Theorem 7 to a p^4-ary Reed-Solomon code of dimension p^3 (relative minimum distance $> 1 - (1/p)$) and an inner array $(p^2, 4)_p$, which is $(1/p)-$biased. Such an array follows from the WCU construction. We can describe it as follows: Its rows are $(x, y, xy, x^2 + cy^2)$, where $x, y \in \mathbb{F}_p$ and c is a non-square. The result is an

$$\frac{2}{p} - \text{ biased } (p^6, 4p^3)_p - \text{ array,}$$

which is better than what results from the WCU construction.

Example 2. In the same style apply Theorem 7 to a p^m-ary Reed-Solomon code of dimension p^{m-1} and an $(1/p)-$biased array $(m^2p^2, m)_p$, whose existence is guaranteed by Theorem 4. We obtain an

$$\frac{2}{p} - \text{ biased } (m^2p^{m+2}, mp^{m-1})_p - \text{ array.}$$

This is much better than a corresponding WCU-array. Theorem 4 with the same bias and the same number of columns would use $m^2p^{2m}/4$ rows.

So far the only ingredients used in our constructions have been Reed-Solomon codes. Next we want to show that algebraic-geometric codes can be used to great advantage. Let us start by pointing out that many important classes of algebraic-geometric codes can be just as efficiently implemented as Reed-Solomon codes. In the next section this is exemplified in the case of the Hermitian codes.

6 Hermitian Codes for the User

We describe how to obtain generator matrices for the Hermitian codes. Consider the field extension $\mathbb{F}_{q^2} \mid \mathbb{F}_q$ and the corresponding trace tr and norm N, where $tr(x) = x + x^q, N(x) = x^{q+1}$. Our codes are defined over $\mathbb{F}_{q^2}$ and have length q^3 (see [24]).

The coordinates are parametrized by the pairs (α, β), where $N(\alpha) = tr(\beta)$. So we need to calculate traces and norms of all elements in the field and to list all these pairs in some order. There are q^3 such pairs.

The general build-up: We construct a $(q^3 - g, q^3)-$ matrix G with entries from $\mathbb{F}_{q^2}$. Here $g = \binom{q}{2}$. The first k rows of G generate the $k-$dimensional Hermitian code. It has parameters

$$[q^3, k, q^3 - k + 1 - g]_{q^2}.$$

The pole-order test: For $n = 0, 1, 2, \dots$ we have to decide if n is a **pole-order** or not. If n is a pole-order we determine its **coordinate vector** (i, j). This is done as follows: Let r be the remainder of n mod q, where $0 \leq r \leq q - 1$ and $-s$ the (negative) remainder of n mod $q + 1$, where $0 \leq s \leq q$. Then n is a pole-order if and only if

$$x = \frac{n - r}{q} \geq \frac{n + s}{q + 1} = y.$$

If $n \geq 2g$, then the pole-order test does not need to be performed. Every such number is a pole-order. If n is a pole-order, then $n = (q + 1)i + qj$, where $i = (x - y)q + r, j = s$. The coordinate vector of n is (i, j).

Constructing the rows of G **:** Let $u_1 = 0, u_2 = q, u_3 = q + 1 \dots$ be the first pole-orders. If u_k has coordinate-vector (i, j), then the entry of row k of G in coordinate (α, β) is $\beta^i \alpha^j$.

We conclude that the use of Hermitian codes requires the usual field arithmetic, just as Reed-Solomon codes.

7 Using Hermitian and Suzuki Codes

Use Theorem 7 with the Hermitian codes as ingredients, $q = p^m$. The codes have parameters

$$[p^{3m}, k, p^{3m} - (k + p^{2m}/2)]_{p^{2m}}.$$

Use as inner arrays the unbiased arrays $(p^{2m}, 2m)_p$. Choose $e \leq m$ and $k \sim p^{3m-e} - p^{2m}/2$. With this choice the resulting array has bias $\epsilon \leq p^{-e}$. As we have an array $(p^{5m}, 2km)_p$ and $\log_p(2km) \sim 3m - e + \log_p(m)$ it follows $f_p(3m - e + \log_p(m), e) \leq 5i$, where $m \geq e$.

Let now e and b be given, where $b \geq 2e$. Determine $m \geq e$ such that $b + e = 3m$ (provided $b + e$ is a multiple of 3). We have seen that $f_p(b, e) \leq$

$5m = \frac{5}{3}(b+e)$, which clearly represents an improvent on Theorem 4 and on the $WCU-$construction. If $b < 2e$, then $f_p(b,e) \leq f_p(2e,e) \leq 5e$, still an improvement upon Theorem 4 when $b \geq \frac{3}{2}e$.

The **Suzuki codes** in characteric 2 (see [12]) have parameters

$$[2^{4f+2}, 2^j, 2^{4f+2} - (2^j + 2^{3f+1})]_{2^{2f+1}}.$$

Use Theorem 7 with an unbiased array as inner array. If $f \geq e$ and $j = 4f - e + 1$ we obtain $\epsilon \leq 2^{-e}$, and hence $f_2(4f - e + 1, e) \leq 6f + 3$. This presupposes $b + e = 4f + 1 > 4e$, hence $b > 3e$.

Theorem 10. *The Hermitian codes show*

$$f_p(b,e) \leq \frac{5}{3}(b+e) \text{ if } b \geq 2e.$$

The Suzuki codes show

$$f_2(b,e) \leq \frac{3}{2}(b+e) + 2 \text{ if } b > 3e.$$

The results of Theorem 10 are superior to all the constructions discussed earlier, for the parameter range when Theorem 10 applies. The strength of Theorem 4 is its universality and simplicity. For $b < e$ it seems to be hard to obtain improvements upon the WCU-construction. Another construction principle for weakly biased arrays, first introduced in [18], uses expander graphs and asymptotically nontrivial families of codes as ingredients. However, this construction seems to work best when k is large with respect to $1/\epsilon$ (b large with respect to e) and it cannot improve upon the results presented above in that parameter range.

We conclude this section with an application of Theorem 7 to Hermitian codes. The p^2-ary Hermitian code of dimension $k \sim p^2/2$ has relative minimum weight $\delta = 1 - 1/p$. The unbiased $(p^2, 2)_p-$array yields an $(1/p)-$biased $(p^5, p^2)_p-$array.

Example 3. For every odd prime p we can produce an $(1/p)-$biased $(p^5, p^2)_p-$array by applying Theorem 7 to a Hermitian code and an unbiased array.

Observe that the WCU construction when applied in the case of p^5 rows and $\epsilon = 1/p$ produces a number of columns of the order of magnitude $p^{3/2}$.

8 Construction of Authentication Schemes

Unconditional authentication was originally introduced by Simmons [22, 23]. An $(n,k)_q-$array is $\epsilon-$**almost strongly universal**$_2$ (ASU$_2$) if each column has bias 0 and for any two different columns c, c' and any entries e, e' the conditional probability $Pr(c_i = e \mid c'_i = e')$ is bounded by ϵ, where the probability refers to a choice of a row i according to the uniform distribution of rows. In the application rows are keys, columns are source states and entries are authentication tags. A composition construction based on codes is used in [4, 2, 3]. In [13] a direct link is established between the WCU construction of weakly biased arrays and ASU_2-arrays. We generalize this construction as follows:

Theorem 11. *If there is an ϵ_0-biased $(n,k)_p-$array then for every $t \leq k$ there is an $\epsilon-ASU_2$ array $(p^t n, p^k)_{p^t}$, where $\epsilon = p^{-t} + \epsilon_0$.*

Proof. Let $\mathcal{C}$ be the linear $[n,k]_p$-code generated by the columns of the ϵ_0-biased array. The columns of the ASU_2-array $\mathcal{A}$ are indexed by $f \in \mathcal{C}$, the rows are indexed by tuples $(i, \alpha_1, \dots, \alpha_t)$, where i is a coordinate of $\mathcal{C}$ and $\alpha_r \in \mathbb{F}_p$. It is easy to see that we can find linear mappings $M_r : \mathcal{C} \longrightarrow \mathcal{C},\ r = 1, 2 \dots, t$ such that every nontrivial $\mathbb{F}_p-$linear combination of the M_r is non-singular. Define the entry of $\mathcal{A}$ in row $(i, \alpha_1, \dots, \alpha_t)$ and column f as $(M_1(f)(i) + \alpha_1, \dots, M_t(f)(i) + \alpha_r)$.

It is obvious that each column of $\mathcal{A}$ is unbiased. Let f, g be different columns and $(\beta_r), (\gamma_r)$ be two entries. Let ν be the number of rows i of the original array such that $M_r(f-g)(i) = \beta_r - \gamma_r$ for all r. We have to show that $\nu/n \leq p^{-t} + \epsilon_0$. This follows from Theorem 2 and the linear independence of the $M_r(f-g)$.

We see that via Theorem 11 essential improvements upon the parameters of weakly biased arrays yield improved authentication$_2$ codes.

Example 4. Continuing from Example 3 we obtain $(p^6, p^{p^2})_p$ arrays, which are $(2/p)-\mathrm{ASU}_2$. Not surprisingly this is better than the constructions from [4, 13] based on Reed-Solomon codes and it reproduces the parameters of the construction from [2] based on Hermitian codes.

Example 5. An application of Theorem 11 to the arrays from Example 2 produces arrays $(m^2 p^{m+3}, p^{mp^{m-1}})_p$, which are $(3/p)-\mathrm{ASU}_2$.

A refinement of the theory of unconditional authentication is introduced in [16]. An $(N,m)_p-$array is $(\delta, t)-$**almost strongly universal** (short $(\delta,t)-$ASU) if for every set $U = U_0 \cup \{u\}$ of t columns and every $a' \in \mathbb{F}_p^{t-1}, x \in \mathbb{F}_p$ the frequencies $\nu_{U_0}(a')$ and $\nu_U(a', x)$ satisfy

$$\mid \nu_U(a', x)/\nu_{U_0}(a') \mid \leq \delta.$$

The idea is to use the same key for t subsequent messages while still bounding the opponent's probability of success. The link between almost independent arrays and $(\delta,t)-$ASU codes has been established in [16] (and is almost obvious):

Theorem 12. *A t-wise $\epsilon-$ dependent array is $(\delta,t)-ASU$, where $\delta = (p^{-t} + \epsilon)/(p^{-(t-1)} - \epsilon)$.*

The following theorem generalizes the method used in [16].

Theorem 13. *Let $f_2(b, lt) \leq a$. Then there is an $(2^{-(l-1)}, t)-ASU$ with 2^l entries, $2^b/(lt) - \log_2(l)$ source bits and a key bits.*

Proof. A BCH-code $[2^j, 2^j - ljt, lt+1]_2$, where $jlt = 2^b$, yields an $lt-$wise $2^{-e}-$biased array $(2^a, 2^j)$. By Theorem 3 this yields an array $(2^a, \frac{1}{l}2^j)_{2^l}$, which is $t-$wise $2^{-lt}-$biased. Apply Theorem 12. We obtain $\delta < 2/(2^l - 1) \sim 2^{-(l-1)}$. The number of rows is still 2^a.

9 Resiliency

A number of interesting applications of the WCU construction are in [16]. They can all be generalized to admit the use of arbitrary weakly biased arrays. We consider the case of almost resilient functions. The construction from [16] is an application of Theorem 1 to check matrices of binary BCH codes. A straightforward generalization is as follows:

Theorem 14. *Assume the following exist:*

- *A systematic* $\epsilon-$*biased* $(2^t, s)_2-$*array, and*
- *a linear code* $[m, m-s, k+1]_2$.

Then there exists a systematic $k-$*wise* $\epsilon-$*dependent* $(2^t, m)_2-$*array*

The proof is similar to the proof for the special case used in [16]. The end product of Theorem 14 allows the construction of a function : $\mathbb{F}_2^m \longrightarrow \mathbb{F}_2^{m-t}$ such that whenever k of the input parameters are fixed the output is close to being unbiased (for details see [16]). Note that the study of almost resilient functions can be motivated from an analysis of the wire-tap channel of type II [20]. A discussion of that aspect is in [26], where the close link to the coding-theoretic and geometric notion of generalized Hamming weights is pointed out.

10 Conclusion

The concepts of sample spaces which are statistically close to being unbiased or independent is fundamental for large areas of computer science and cryptology. The best known constructions all yield very similar parameters. The various constructions from [1] excel by their simplicity and universality, whereas the Weil-Carlitz-Uchiyama construction yields slightly better parameters. In this paper we used several new coding-theoretic construction procedures to obtain essential improvements for vast parameter ranges. These improvements can already be obtained by restricting the ingredients to Reed-Solomon codes. Algebraic-geometric codes produce further improvements in suitable parameter ranges. We pointed out that Hermitian codes, a particularly useful class of AG codes, are just as efficiently computable as Reed-Solomon codes.

In the applications we concentrated on universal hashing, unconditional authentication and almost resilient functions. A large number of applications are documented in the literature. It is expected that more applications will be discovered.

References

1. Alon, N., Goldreich, O., Håstad, J., Peralta, R.: Simple constructions of almost k-wise independent random variables, Random Structures and Algorithms **3** (1992), 289-304, preliminary version: Symposium 31st FOCS 1990, 544-553
2. Bierbrauer, J.: Universal hashing and geometric codes, Designs, Codes and Cryptography **11** (1997), 207-221
3. Bierbrauer, J.: Authentication via algebraic-geometric codes, in: Recent Progress in Geometry, Supplemento ai Rendiconti del Circolo Matematico di Palermo **51** (1998), 139-152
4. Bierbrauer, J., Johansson, T., Kabatiansky, G., Smeets, B.: On families of hash functions via geometric codes and concatenation, Proceedings CRYPTO 93, Lecture Notes in Computer Science **773** (1994), 331-342
5. Bierbrauer, J., Schellwat, H.: Weakly biased arrays, almost independent arrays and error-correcting codes, submitted for publication in the Proceedings of AMS-DIMACS.
6. Boyar, J., Brassard, G., Peralta, R.: Subquadratic zero-knowledge, JACM **42** (1995), 1169-1193
7. Brassard, G., Crépeau, C., Santha,M.: Oblivious transfers and intersecting codes, IEEE Transactions on Information Theory **42** (1996), 1769-1780
8. Carlitz, L., Uchiyama, S.: Bounds for exponential sums, Duke Mathematical Journal **24** (1957), 37-41
9. Cohen, G. D., Zémor, G.: Intersecting codes and independent families, IEEE Transactions on Information Theory **40** (1994), 1872-1881
10. Gal, A.: A characterization of span program size and improved lower bounds for monotone span programs, Proceedings 13^{th} Symposium of the Theory of Computing (1998), 429-437
11. Carter, J. L., Wegman, M. N.: Universal Classes of Hash Functions, J.Computer and System Sci. **18** (1979), 143-154
12. Hansen, J. P., Stichtenoth, H.: Group codes on certain algebraic curves with many rational points, AAECC **1** (1990), 67-77
13. Helleseth, T., Johansson, T.: Universal hash functions from exponential sums over finite fields and Galois rings, Lecture Notes in Computer Science **1109** (1996), 31-44 (CRYPTO 96)
14. Justesen, J.: A class of asymptotically good algebraic codes, IEEE Transactions on Information Theory **18** (1972), 652-656
15. Katsman, G. L., Tsfasman, M. A., Vladut, S. G.: Modular curves and codes with a polynomial construction, IEEE Transaction on Information Theory **30** (1984), 353-355
16. Kurosawa, K., Johansson, T., Stinson, D.: Almost k-wise independent sample spaces and their cryptologic applications, Lecture Notes in Computer Science **1233** (1997), 409-421 (Advances in Cryptology, Eurocrypt 97)
17. Lu, C. J.: Improved pseudorandom generators for combinatorial rectangles, Proceedings of the 25th International Colloquium on Automata, Languages and Programming (1998), 223-234
18. Naor, J., Naor, M.: Small-bias probability spaces: efficient constructions and applications, SIAM Journal on Computing **22** (1993), 838-856, preliminary version: Proceedings STOC 1990, 213-223
19. Naor, M., Reingold, O.: On the construction of pseudo-random permutations: Luby-Rackoff revisited, Proceedings STOC **29** (1997), 189-199

20. Ozarow, L. H., Wyner, A. D.: Wire-Tap Channel II, AT&T Bell Laboratories Technical Journal **63** (1984), 2135-2157
21. Shen, B. Z.: A Justesen construction of binary concatenated codes that asymptotically meet the Zyablov bound for low rate, IEEE Transactions on Information Theory **39** (1993), 239-242
22. Simmons, G. J.: A game theory model of digital message authentication, Congressus Numerantium **34** (1992), 413-424
23. Simmons, G. J.: Authentication theory/coding theory, in: Advances in Cryptology, Proceedings of Crypto 84, Lecture Notes in Computer Science **196** (1985), 411-431
24. Stichtenoth, H.: Algebraic function fields and codes, Springer 1993.
25. Wegman, M. N., Carter, J. L.: New Hash Functions and Their Use in Authentication and Set Equality, J.Computer and System Sci. **22** (1981), 265-279
26. Wei, V. K.: Generalized Hamming weights for linear codes, IEEE Transactions on Information Theory **37** (1991), 1412-1418
27. Zyablov, V. V.: An estimate of the complexity of constructing binary linear cascade codes, Problems in Information transmission **7** (1971), 3-10